Projective Splitting with Forward Steps only Requires Continuity

Patrick R. Johnstone[*] Jonathan Eckstein[*]

September 17, 2018

Abstract

A recent innovation in projective splitting algorithms for monotone operator inclusions has been the development of a procedure using two forward steps instead of the customary proximal steps for operators that are Lipschitz continuous. This paper shows that the Lipschitz assumption is unnecessary when the forward steps are performed in finite-dimensional spaces: a backtracking linesearch yields a convergent algorithm for operators that are merely continuous with full domain.

1 Introduction

For a collection of real Hilbert spaces $\{\mathcal{H}_i\}_{i=0}^n$, consider the problem of finding $z \in \mathcal{H}_0$ such that

$$0 \in \sum_{i=1}^{n} G_i^* T_i(G_i z), \tag{1}$$

where $G_i : \mathcal{H}_0 \to \mathcal{H}_i$ are linear and bounded operators, $T_i : \mathcal{H}_i \to 2^{\mathcal{H}_i}$ are maximal monotone operators. We suppose that T_i is continuous with $\operatorname{dom}(T_i) = \mathcal{H}_i$ for each i in some subset $\mathcal{I}_{\mathrm{F}} \subseteq \{1, \ldots, n\}$. A key special case of (1) is

$$\min_{x \in \mathcal{H}_0} \sum_{i=1}^{n} f_i(G_i x), \tag{2}$$

where every $f_i : \mathcal{H}_i \to \mathbb{R}$ is closed, proper and convex, with some subset of the functions also being Fréchet differentiable everywhere. Under appropriate constraint qualifications, (1) and (2) are equivalent. Problem (2) arises in a host of applications such as machine learning, signal and image processing, inverse problems, and computer vision; see [4, 6, 7] for some examples.

[*] Department of Management Science and Information Systems, Rutgers Business School Newark and New Brunswick, Rutgers University

A relatively recently proposed class of operator splitting algorithms which can solve (1) is *projective splitting*. It originated with [10] and was then generalized to more than two operators in [11]. The related algorithm in [1] introduced a technique for handling compositions of linear and monotone operators, and [5] proposed an extension to "block-iterative" and asynchronous operation — block-iterative operation meaning that only a subset of the operators making up the problem need to be considered at each iteration (this approach may be called "incremental" in the optimization literature). A restricted and simplified version of this framework appears in [9]. Our recent work in [14] incorporated forward steps into the projective splitting framework, in place of the customary proximal (backward) steps, for any Lipschitz continuous operators, and introduced backtracking and adaptive stepsize rules; see also [21]. The even more recent work [13] derived convergence rates for the method under various conditions.

In general, projective splitting offers unprecedented levels of flexibility compared with previous operator splitting algorithms such as [18, 16, 8, 22]. The framework can be applied to arbitary sums of maximal monotone operators, the stepsizes can vary by operator and by iteration, compositions with linear operators can be handled, and block-iterative asynchronous implementations have been demonstrated. Furthermore the number of times each operator is processed does not need to be equal (either exactly or approximately).

In [14], we showed that it is possible for projective splitting to process Lipschitz-continuous operators using a pair of forward steps rather than the customary proximal step. In general, the stepsize must be bounded by the inverse of the Lipschitz constant, but a backtracking linesearch procedure is available when this constant is unknown. See also [21] for a similar approach to using forward steps in a more restrictive projective splitting context, without backtracking.

The purpose of this work is to show that this Lipschitz assumption is unnecessary. It demonstrates that, when the Hilbert spaces $\mathcal{H}_i$ in (1) are finite dimensional for $i \in \mathcal{I}_{\text{F}}$, the two-forward-step procedure with backtracking linesearch yields weak convergence to a solution assuming only simple continuity and full domain of the operators T_i.[1] A new argument is required beyond those in [14] since the stepsizes resulting from the backtracking linesearch are no longer guaranteed to be bounded away from 0.

Theoretically, this result aligns projective splitting with two related monotone operator splitting methods which utilize two forward steps per iteration and only require continuity in finite dimension. These are Tseng's forward-backward-forward method [22] and the extragradient method [15, 12, 3]. These methods apply to special cases of Problem (1) with $n = 2$, $\mathcal{I}_{\text{F}} = \{1\}$, and $G_1 = G_2 = I$; the extragradient method also restricts T_2 to be the normal cone map of a closed convex set. While the original extragradient method [15] was applied to variational inequalities under Lipschitz continuity, it was extended in [12] to include a backtracking linesearch that works under continuity alone and in [3] to solve monotone inclusions. In fact, the algorithm of [12, 3] for just one operator is almost the special case of projective splitting applied to Problem (1) with one operator and $\mathcal{I}_{\text{F}} = \{1\}$, the only difference being a stricter criterion to terminate the backtracking linesearch. Tseng's method can also be

[1] We still speak of weak convergence because the spaces $\mathcal{H}_i$ may be infinite dimensional for $i \notin \mathcal{I}_{\text{F}}$. If $\mathcal{H}_i$ is infinite dimensional for $i \in \mathcal{I}_{\text{F}}$, we can instead require T_i to be Cauchy continuous for all bounded sequences.

connected to projective splitting with one operator and $\mathcal{I}_{\mathrm{B}} = \{1\}$ following the arguments of [20, Section 5.1].

All of these methods can be viewed in contrast with the classical forward-backward splitting algorithm [18]. This method utlizes a single forward step at each iteration but requires a cocoercivity assumption which is in general stricter than Lipschitz continuity. Also disadvantageous is that the choice of stepsize depends on knowledge of the cocoercivity constant and no backtracking linesearch is known to be available. Progress was made in a very recent paper [17] which modified the forward-backward method so that it can be applied to (locally) Lipschitz continuous operators with backtracking for unknown Lipschitz constant. The locally Lipschitz continuous assumption is stronger than the mere continuity assumption considered here and in [22, 12, 3], and for known Lipschitz constant the stepsize constraint is more restrictive.

The rest of the paper is organized as follows: In Section 2, we present notation and some basic background results. In Section 3, we precisely state the projective splitting algorithm and our assumptions, and collect some necessary results from [14]. Finally section 4 proves the main result.

2 Preliminaries and Notation

As in [14], will work with a slight restriction of problem (1), namely

$$0 \in \sum_{i=1}^{n-1} G_i^* T_i(G_i z) + T_n(z). \tag{3}$$

In terms of problem (1), we are simply requiring that G_n be the identity operator and thus that $\mathcal{H}_n = \mathcal{H}_0$. This is not much of a restriction in practice, since one could redefine the last operator as $T_n \leftarrow G_n^* \circ T_n \circ G_n$, or one could simply append a new operator T_n with $T_n(z) = \{0\}$ everywhere.

We will use a boldface $\mathbf{w} = (w_1, \ldots, w_{n-1})$ for elements of $\mathcal{H}_1 \times \ldots \times \mathcal{H}_{n-1}$. To ease the presentation, we use the following notation throughout, where I denotes the identity operator:

$$G_n : \mathcal{H}_n \to \mathcal{H}_n \triangleq I \qquad\qquad (\forall\, k \in \mathbb{N}) \quad w_n^k \triangleq -\textstyle\sum_{i=1}^{n-1} G_i^* w_i^k. \tag{4}$$

For any maximal monotone operator A we will use the notation

$$\operatorname{prox}_{\rho A} = (I + \rho A)^{-1}$$

for any scalar $\rho > 0$ to denote the *proximal operator*, also known as the backward or implicit step with respect to A. This means that

$$x = \operatorname{prox}_{\rho A}(a) \quad \Longrightarrow \quad \exists y \in Ax : x + \rho y = a.$$

The x and y satisfying this relation are unique. Furthermore, $\operatorname{prox}_{\rho A}$ is defined everywhere and $\operatorname{range}(\operatorname{prox}_A) = \operatorname{dom}(A)$ [2, Prop. 23.2].

By continuity, we mean in the strong topology defined in terms of the norm of the given Hilbert space. That is, for all $g_0 \in \mathcal{H}_i$ and $\epsilon > 0$, there exists $\delta(g_0, \epsilon)$ s.t. whenever $\|g_0 - g\| \leq \delta(g_0, \epsilon)$, $\|T_i(g_0) - T_i(g)\| \leq \epsilon$. Uniform continuity means that the constant is independent of g_0, i.e. the above statement holds with $\delta(g_0, \epsilon) = \delta(\epsilon)$.

We use the standard "$\rightharpoonup$" notation to denote weak convergence, which is of course equivalent to ordinary convergence in finite dimensional settings.

3 Algorithm, Principal Assumptions, and Preliminary Analysis

Algorithm 1 presents the algorithm analyzed in this paper. It is essentially the block-iterative and potentially asynchronous projective splitting algorithm as in [14], but directly incorporating a backtracking linesearch procedure.

Let $\boldsymbol{\mathcal{H}} = \mathcal{H}_0 \times \mathcal{H}_1 \times \cdots \times \mathcal{H}_{n-1}$ and $\mathcal{H}_n = \mathcal{H}_0$. The algorithm produces a sequence of iterates denoted by $p^k = (z^k, w_1^k, \ldots, w_{n-1}^k) \in \boldsymbol{\mathcal{H}}$. Define the *extended solution set* or *Kuhn-Tucker set* of (3) to be

$$\mathcal{S} = \left\{ (z, \mathbf{w}) \in \boldsymbol{\mathcal{H}} \,\middle|\, w_i \in T_i(G_i z),\ i = 1, \ldots, n-1, -\sum_{i=1}^{n-1} G_i^* w_i \in T_n(z) \right\}. \tag{5}$$

Clearly $z \in \mathcal{H}_0$ solves (3) if and only if there exists $\mathbf{w} \in \mathcal{H}_1 \times \cdots \times \mathcal{H}_{n-1}$ such that $(z, \mathbf{w}) \in \mathcal{S}$.

Algorithm 1 is a special case of a general seperator-projector method for finding a point in a closed and convex set. At each iteration the method constructs an affine function $\varphi_k : \mathcal{H}^n \to \mathbb{R}$ which separates the current point from the target set $\mathcal{S}$ defined in (5). In other words, if p^k is the current point in $\boldsymbol{\mathcal{H}}$ generated by the algorithm, $\varphi_k(p^k) > 0$, and $\varphi_k(p) \leq 0$ for all $p \in \mathcal{S}$. The next point is then the projection of p^k onto the hyperplane $\{p : \varphi_k(p) = 0\}$, subject to a relaxation factor β_k. What makes projective splitting an operator splitting method is that the hyperplane is constructed through individual calculations on each operator T_i, either prox calculations or forward steps.

The hyperplane is defined in terms of the following affine function:

$$\varphi_k(z, w_1, \ldots, w_{n-1}) = \sum_{i=1}^{n-1} \langle z - x_i^k, y_i^k - w_i \rangle + \left\langle z - x_i^n, y_i^n + \sum_{i=1}^{n-1} w_i \right\rangle. \tag{6}$$

See [14, Lemma 4] for the relevent properties of φ_k. As in [14], we use the following inner product and norm for $\boldsymbol{\mathcal{H}}$, for an arbitrary scalar $\gamma > 0$:

$$\begin{gathered} \left\langle (z^1, \mathbf{w}^1), (z^2, \mathbf{w}^2) \right\rangle_\gamma = \gamma \langle z^1, z^2 \rangle + \textstyle\sum_{i=1}^{n-1} \langle w_i^1, w_i^2 \rangle \\ \|(z, \mathbf{w})\|_\gamma^2 = \langle (z, \mathbf{w}), (z, \mathbf{w}) \rangle_\gamma . \end{gathered}$$

Note that with this inner product it was shown in [14, Lemma 4] that

$$\nabla \varphi_k = \left(\frac{1}{\gamma} \left(\sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k \right),\ x_1^k - G_1 x_n^k, x_2^k - G_2 x_n^k, \ldots, x_{n-1}^k - G_{n-1} x_n^k \right). \tag{7}$$

The scalar $\gamma > 0$ controls the relative emphasis on the primal and dual variables in the projection update in lines 37-38.

Algorithm 1 has the following parameters:

- For each iteration $k \geq 1$, a subset $I_k \subseteq \{1, \ldots, n\}$.
- For each $k \geq 1$ and $i = 1, \ldots, n$, a positive scalar stepsize ρ_i^k. For $i \in \mathcal{I}_{\text{F}}$, ρ_i^k is the initial stepsize tried in the backtracking linesearch while $\hat{\rho}_i^k$ is the accepted stepsize.
- A constant $\nu \in (0, 1)$ controlling how much the stepsize is decreased at each iteration of the backtracking linesearch.
- For each iteration $k \geq 1$ and $i = 1, \ldots, n$, a delayed iteration index $d(i, k) \in \{1, \ldots, k\}$ which allows the subproblem calculations to use outdated information.
- For each iteration $k \geq 1$, an overrelaxation parameter $\beta_k \in [\underline{\beta}, \overline{\beta}]$ for some constants $0 < \underline{\beta} \leq \overline{\beta} < 2$. see [14] for more details.
- Sequences of errors $\{e_i^k\}_{k \geq 1}$ for $i \in \mathcal{I}_{\text{B}}$, allowing us to model inexact computation of the proximal steps.

There are many ways in which Algorithm 1 could be implemented in various parallel computing environments. We refer to [14] for a more thorough discussion.

Our main assumptions regarding (3) are as follows:

Assumption 1. *Problem (3) conforms to the following:*

1. *$\mathcal{H}_0 = \mathcal{H}_n$ and $\mathcal{H}_1, \ldots, \mathcal{H}_{n-1}$ are real Hilbert spaces.*
2. *For $i = 1, \ldots, n$ the operators $T_i : \mathcal{H}_i \to 2^{\mathcal{H}_i}$ are monotone.*
3. *For all i in some subset $\mathcal{I}_{\text{F}} \subseteq \{1, \ldots, n\}$, $\mathcal{H}_i$ is finite-dimensional, the operator T_i is continuous with respect to the metric induced by $\|\cdot\|$ (and thus single-valued), and $dom(T_i) = \mathcal{H}_i$.*
4. *For $i \in \mathcal{I}_{\text{B}} \triangleq \{1, \ldots, n\} \backslash \mathcal{I}_{\text{F}}$, the operator T_i is maximal and the map $prox_{\rho T_i} : \mathcal{H}_i \to \mathcal{H}_i$ can be computed to within the error tolerance specified below in Assumption 3.*
5. *Each $G_i : \mathcal{H}_0 \to \mathcal{H}_i$ for $i = 1, \ldots, n-1$ is linear and bounded.*
6. *The solution set $\mathcal{S}$ defined in (5) is nonempty.*

Our assumptions regarding the parameters of Algorithm 1 are as follows, and are the same as used in [5, 9, 14].

Assumption 2. *For Algorithm 1, assume:*

1. *For some fixed integer $M \geq 1$, each index i in $1, \ldots, n$ is in I_k at least once every M iterations, that is, $\bigcup_{k=j}^{j+M-1} I_k = \{1, \ldots, n\}$ for all $i = 1, \ldots, n$ and $j \geq 1$.*
2. *For some fixed integer $D \geq 0$, we have $k - d(i, k) \leq D$ for all i, k with $i \in I_k$.*

Algorithm 1: Asynchronous algorithm for solving (3).

```
Input: $(z^1, \mathbf{w}^1) \in \boldsymbol{\mathcal{H}}$, $(x_i^0, y_i^0) \in \mathcal{H}_i^2$ for $i = 1, \ldots, n$, $0 < \underline{\beta} \leq \overline{\beta} < 2$, $\gamma > 0$, $\nu \in (0, 1)$.
for $k = 1, 2, \ldots$ do
    for $i = 1, 2, \ldots, n$ do
        if $i \in I_k$ then
            if $i \in \mathcal{I}_{\mathrm{B}}$ then
                $a = G_i z^{d(i,k)} + \rho_i^{d(i,k)} w_i^{d(i,k)} + e_i^k$
                $x_i^k = \mathrm{prox}_{\rho_i^{d(i,k)} T_i}(a)$
                $y_i^k = (\rho_i^{d(i,k)})^{-1} \left(a - x_i^k\right)$
            else
                $\rho_i^{(1,k)} \leftarrow \rho_i^{d(i,k)}$
                $\theta_i^k = G_i z^{d(i,k)}$
                $\zeta_i^k = T_i \theta_i^k$
                if $\zeta_i^k = w_i^{d(i,k)}$ then
                    $\hat{\rho}_i^{d(i,k)} \leftarrow \rho_i^{(j,k)},\ x_i^k \leftarrow \theta_i^k,\ y_i^k \leftarrow \zeta_i^k$
                else
                    for $j = 1, 2, \ldots$ do
                        $\tilde{x}_i^{(j,k)} = \theta_i^k - \rho_i^{(j,k)}(\zeta_i^k - w_i^{d(i,k)})$
                        $\tilde{y}_i^{(j,k)} = T_i \tilde{x}_i^{(j,k)}$
                        if $\Delta\|\theta_i^k - \tilde{x}_i^{(j,k)}\|^2 - \langle \theta_i^k - \tilde{x}_i^{(j,k)}, \tilde{y}_i^{(j,k)} - w_i^{d(i,k)} \rangle \leq 0$ then
                            $\hat{\rho}_i^{d(i,k)} \leftarrow \rho_i^{(j,k)},\ x_i^k \leftarrow \tilde{x}_i^{(j,k)},\ y_i^k \leftarrow \tilde{y}_i^{(j,k)}$
                            break
                        $\rho_i^{(j+1,k)} = \nu \rho_i^{(j,k)}$
                    $\hat{\rho}_i^{d(i,k)} \leftarrow \rho_i^{(j,k)}, x_i^k \leftarrow \tilde{x}_i^{(j,k)}, y_i^k \leftarrow \tilde{y}_i^{(j,k)}$
        else
            $(x_i^k, y_i^k) = (x_i^{k-1}, y_i^{k-1})$
    $u_i^k = x_i^k - G_i x_n^k, \quad i = 1, \ldots, n-1,$
    $v^k = \sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k$
    $\pi_k = \|u^k\|^2 + \gamma^{-1}\|v^k\|^2$
    if $\pi_k > 0$ then
        Choose some $\beta_k \in [\underline{\beta}, \overline{\beta}]$
        $\varphi_k(p^k) = \langle z^k, v^k \rangle + \sum_{i=1}^{n-1} \langle w_i^k, u_i^k \rangle - \sum_{i=1}^{n} \langle x_i^k, y_i^k \rangle$
        $\alpha_k = \frac{\beta_k}{\pi_k} \max\left\{0, \varphi_k(p^k)\right\}$
    else
        if $\cup_{j=1}^k I_j = \{1, \ldots, n\}$ then
            return $z^{k+1} \leftarrow x_n^k, w_1^{k+1} \leftarrow y_1^k, \ldots, w_{n-1}^{k+1} \leftarrow y_{n-1}^k$
        else
            $\alpha_k = 0$
    $z^{k+1} = z^k - \gamma^{-1} \alpha_k v^k$
    $w_i^{k+1} = w_i^k - \alpha_k u_i^k, \quad i = 1, \ldots, n-1,$
    $w_n^{k+1} = -\sum_{i=1}^{n-1} G_i^* w_i^{k+1}$
```

We also use the following additional notation from [9]: for all i and k, define

$$S(i,k) = \{j \in \mathbb{N} : j \leq k, i \in I_j\} \qquad s(i,k) = \begin{cases} \max S(i,k), & \text{when } S(i,k) \neq \emptyset \\ 0, & \text{otherwise.} \end{cases}$$

Essentially, $s(i,k)$ is the most recent iteration up to and including k in which the index-i information in the separator was updated. Assumption 2 ensures that $0 \leq k - s(i,k) < M$. For all $i = 1, \ldots, n$ and iterations k, also define $l(i,k) = d(i, s(i,k))$, the iteration in which the algorithm generated the information $z^{l(i,k)}$ and $w_i^{l(i,k)}$ used to compute the current point (x_i^k, y_i^k). Regarding initialization, we set $d(i,0) = 0$; note that the initial points (x_i^0, y_i^0) are arbitrary. We formalize the use of $l(i,k)$ in the following Lemma from [14]:

Lemma 1. *Suppose Assumption 2(1) holds. For all iterations $k \geq M$ if Algorithm 1 has not already terminated, the updates can be written as*

$$(\forall i \in \mathcal{I}_\text{B}) \qquad x_i^k + \rho_i^{l(i,k)} y_i^k = G_i z^{l(i,k)} + \rho_i^{l(i,k)} w_i^{l(i,k)} + e_i^{s(i,k)} \qquad y_i^k \in T_i x_i^k \tag{8}$$

$$(\forall i \in \mathcal{I}_\text{F}) \qquad x_i^k = G_i z^{l(i,k)} - \hat{\rho}_i^{l(i,k)} \big(T_i G_i z^{l(i,k)} - w_i^{l(i,k)}\big) \qquad y_i^k = T_i x_i^k. \tag{9}$$

Proof. Follows directly from [14, Lemma 6], in view of how $\hat{\rho}_i^k$ is calculated in Algorithm 1. □

Since Algorithm 1 is a projection method, it satisfies the following lemma, identical to [14, Lemmas 2 and 6]:

Lemma 2. *Suppose assumptions 1 and 2(1) hold. Then for Algorithm 1*

1. *The sequence $\{p^k\}$ generated by Algorithm 1 is bounded.*
2. *If Algorithm 1 runs indefinitely, then $\|p^k - p^{k+1}\| \to 0$.*
3. *Lines 37 and 38 may be written as*

$$p^{k+1} = p^k - \frac{\beta_k \max\{\varphi_k(p^k), 0\}}{\|\nabla \varphi_k\|_\gamma^2} \nabla \varphi_k.$$

The assumptions regarding the proximal operator evaluation errors are identical to those in [9]:

Assumption 3. *The error sequences $\{\|e_i^k\|\}$ are bounded for all $i \in \mathcal{I}_\text{B}$. For some σ with $0 \leq \sigma < 1$ the following hold for all $k \geq 1$ such that Algorithm 1 has not yet terminated:*

$$(\forall i \in \mathcal{I}_\text{B}) \qquad \langle G_i z^{l(i,k)} - x_i^k, e_i^{s(i,k)} \rangle \geq -\sigma \|G_i z^{l(i,k)} - x_i^k\|^2 \tag{10}$$

$$(\forall i \in \mathcal{I}_\text{B}) \qquad \langle e_i^{s(i,k)}, y_i^k - w_i^{l(i,k)} \rangle \leq \rho_i^{l(i,k)} \sigma \|y_i^k - w_i^{l(i,k)}\|^2. \tag{11}$$

The stepsize assumptions differ from [14, 13] for $i \in \mathcal{I}_\text{F}$ in that we no longer assume Lipschitz continuity nor that the stepsizes are bounded by the inverse of the Lipschitz constant. However, the initial trial stepsize for the backtracking linesearch at each iteration is assumed to be bounded from above and below:

Assumption 4. *In Algorithm 1,*

$$\underline{\rho} \triangleq \min_{i=1,\ldots,n} \left\{ \inf_{k \geq 1} \rho_i^k \right\} > 0 \qquad \overline{\rho} \triangleq \max_{i=1,\ldots,n} \left\{ \sup_{k \geq 1} \rho_i^k \right\} < \infty.$$

4 Weak Convergence to a Solution

Lemma 3. *Suppose assumptions 1–4 hold. Then for all $k \in \mathbb{N}$ and $i \in I_k$ such that Algorithm 1 has not yet terminated, the backtracking linesearch on lines 9–22 terminates in a finite number of iterations.*

Proof. For $k \geq M$, consider any $i \in \mathcal{I}_{\mathrm{F}} \cap I_k$ and assume that $T_i G_i z^{l(i,k)} \neq w_i^{l(i,k)}$, since backtracking otherwise terminates immediately at line 13. Using the definitions of $s(i,k)$ and $l(i,k)$ and some algebraic manipulation, the condition for terminating the backtracking linesearch given on line 18 may be written as:

$$\frac{\langle G_i z^{l(i,k)} - \tilde{x}_i^{(j,s(i,k))}, \tilde{y}_i^{(j,s(i,k))} - w_i^{l(i,k)} \rangle}{\|G_i z^{l(i,k)} - \tilde{x}_i^{(j,s(i,k))}\|^2} \geq \Delta. \tag{12}$$

For brevity, let $\rho = \rho_i^{(j,s(i,k))} > 0$. Using lines 10, 11, 16, and 17, the left-hand side of (12) may be written

$$\frac{\Big\langle T_i G_i z^{l(i,k)} - w_i^{l(i,k)}, T_i\big(G_i z^{l(i,k)} - \rho(T_i G_i z^{l(i,k)} - w_i^{l(i,k)})\big) - w_i^{l(i,k)} \Big\rangle}{\rho \|T_i G_i z^{l(i,k)} - w_i^{l(i,k)}\|^2}. \tag{13}$$

The numerator of this fraction may be expressed as

$$\Big\langle T_i G_i z^{l(i,k)} - w_i^{l(i,k)}, T_i\big(G_i z^{l(i,k)} - \rho(T_i G_i z^{l(i,k)} - w_i^{l(i,k)})\big) - T_i G_i z^{l(i,k)} \Big\rangle \\ + \|T_i G_i z^{l(i,k)} - w_i^{l(i,k)}\|^2.$$

Substituting this expression into (13) and applying the Cauchy-Schwarz inequality to the inner product yields that the left-hand size of (12) is lower bounded by

$$\frac{\|T_i G_i z^{l(i,k)} - w_i^{l(i,k)}\| - \big\|T_i(G_i z^{l(i,k)} - \rho\big(T_i G_i z^{l(i,k)} - w_i^{l(i,k)})\big) - T_i G_i z^{l(i,k)}\big\|}{\rho \|T_i G_i z^{l(i,k)} - w_i^{l(i,k)}\|.} \tag{14}$$

The continuity of T_i implies that the above expression tends to $+\infty$ as $\rho \to 0$. Since $\rho_j^{(j,k)}$ decreases geometrically to 0 with j on line 21, it follows that (12) must eventually hold. □

Lemma 4. *Suppose assumptions 1–4 hold, Algorithm 1 produces an infinite sequence of iterates, and both*

1. $G_i z^{l(i,k)} - x_i^k \to 0$ *for all* $i = 1, \ldots, n$
2. $y_i^k - w_i^{l(i,k)} \to 0$ *for all* $i = 1 \ldots, n$.

Then the sequence $\{(z^k, \mathbf{w}^k)\}$ generated by Algorithm 1 converges weakly to some point $(\bar{z}, \overline{\mathbf{w}})$ in the extended solution set $\mathcal{S}$ of (3) *defined in* (5). *Furthermore, $x_i^k \rightharpoonup G_i \bar{z}$ and $y_i^k \rightharpoonup \overline{w}_i$ for all $i = 1, \ldots, n-1$, $x_n^k \rightharpoonup \bar{z}$, and $y_n^k \rightharpoonup -\sum_{i=1}^{n-1} G_i^* \overline{w}_i$.*

Proof. First, note that $w_i^{l(i,k)} - w_i^k \to 0$ for all $i = 1, \ldots, n$ and $z^{l(i,k)} - z^k \to 0$ [14, Lemma 9]. Combining $z^k - z^{l(i,k)} \to 0$ with point (1) and the fact that G_i is bounded, we obtain that $G_i z^k - x_i^k \to 0$ for $i = 1, \ldots, n$. Similarly, combining $w_i^{l(i,k)} - w_i^k \to 0$ with point (2) we have $y_i^k - w_i^k \to 0$. The proof is now identical to part 3 of the proof of [14, Theorem 1]. □

Before commencing with the final two technical lemmas, we need two definitions. Define $\phi_k = \varphi_k(p^k)$ and

$$(\forall i = 1, \ldots, n) \quad \psi_{ik} \triangleq \langle G_i z^{l(i,k)} - x_i^k, y_i^k - w_i^{l(i,k)} \rangle \qquad\qquad \psi_k \triangleq \textstyle\sum_{i=1}^n \psi_{ik}. \tag{15}$$

Lemma 5. *Suppose assumptions 1–4 hold and that Algorithm 1 produces an infinite sequence of iterates with $\{x_i^k\}$ and $\{y_i^k\}$ being bounded. Then, for all $i = 1, \ldots, n$, it holds that $G_i z^{l(i,k)} - x_i^k \to 0$.*

Proof. Using (7)

$$\|\nabla\varphi_k\|_\gamma^2 = \gamma^{-1} \left\|\textstyle\sum_{i=1}^{n-1} G_i^* y_i^k + y_n^k\right\|^2 + \textstyle\sum_{i=1}^{n-1} \|x_i^k - G_i x_n^k\|^2. \tag{16}$$

By assumption, $\{x_i^k\}$ and $\{y_i^k\}$ are bounded sequences, therefore $\{\|\nabla\varphi_k\|_\gamma\}$ is bounded; let $\xi_1 > 0$ be some bound on this sequence. Next, we will establish that there exists some $\xi_2 > 0$ such that

$$\psi_k \geq \xi_2 \textstyle\sum_{i=1}^n \|G_i z^{l(i,k)} - x_i^k\|^2. \tag{17}$$

The proof resembles that of [14, Lemma 12]: since the backtracking linesearch terminates in a finite number of iterations, we must have

$$\langle G_i z^{l(i,k)} - x_i^k, y_i^k - w_i^{l(i,k)} \rangle \geq \Delta \|G_i z^{l(i,k)} - x_i^k\|^2 \tag{18}$$

for every $k \in \mathbb{N}$ and $i \in \mathcal{I}_\text{F}$. Terms in $\mathcal{I}_\text{B}$ are treated as before in [14, Lemma 12]: specifically, for all $i \in \mathcal{I}_\text{B}$,

$$\begin{aligned}
\psi_{ik} &= \left\langle G_i z^{l(i,k)} - x_i^k, y_i^k - w_i^{l(i,k)} \right\rangle \\
&\overset{\text{(a)}}{=} \left\langle G_i z^{l(i,k)} - x_i^k, \left(\rho_i^{l(i,k)}\right)^{-1} \left(G_i z^{l(i,k)} - x_i^k + e_i^{s(i,k)}\right) \right\rangle \\
&= \left(\rho_i^{l(i,k)}\right)^{-1} \|G_i z^{l(i,k)} - x_i^k\|^2 + \left(\rho_i^{l(i,k)}\right)^{-1} \left\langle G_i z^{l(i,k)} - x_i^k, e_i^{s(i,k)} \right\rangle \\
&\overset{\text{(b)}}{\geq} (1-\sigma)\left(\rho_i^{l(i,k)}\right)^{-1} \|G_i z^{l(i,k)} - x_i^k\|^2.
\end{aligned} \tag{19}$$

In the above derivation, (a) follows by substitution of (8) and (b) is justified by using (10) in Assumption 3. Combining (18) and (19) yields

$$\psi_k \geq (1-\sigma)\overline{\rho}^{-1} \sum_{i \in \mathcal{I}_\text{B}} \|G_i z^{l(i,k)} - x_i^k\|^2 + \Delta \sum_{i \in \mathcal{I}_\text{F}} \|G_i z^{l(i,k)} - x_i^k\|^2, \tag{20}$$

which yields (17) with $\xi_2 = \min\{(1-\sigma)\overline{\rho}^{-1}, \Delta\} > 0$.

We now proceed as in as in part 1 of the proof of [14, Theorem 1]: first, Lemma 2(3) states that the updates on lines 37–38 can be written as

$$p^{k+1} = p^k - \frac{\beta_k \max\{\phi_k, 0\}}{\|\nabla\varphi_k\|_\gamma^2}\nabla\varphi_k.$$

Lemma 2(2) guarantees that $p^k - p^{k+1} \to 0$, so it follows that

$$0 = \lim_{k\to\infty}\|p^{k+1} - p^k\|_\gamma = \lim_{k\to\infty}\frac{\beta_k \max\{\phi_k, 0\}}{\|\nabla\varphi_k\|_\gamma} \geq \frac{\underline{\beta}\limsup_{k\to\infty}\max\{\phi_k, 0\}}{\sqrt{\xi_1}}.$$

Therefore, $\limsup_{k\to\infty}\phi_k \leq 0$. Since [14, Lemma 10] states that $\phi_k - \psi_k \to 0$, it follows that $\limsup_{k\to\infty}\psi_k \leq 0$. With (a) following from (17), we next obtain

$$\begin{aligned} 0 \geq \limsup_{k\to\infty}\psi_k &\overset{\text{(a)}}{\geq} \xi_2 \limsup_k \textstyle\sum_{i=1}^n \|G_i z^{l(i,k)} - x_i^k\|^2 \\ &\geq \xi_2 \liminf_k \textstyle\sum_{i=1}^n \|G_i z^{l(i,k)} - x_i^k\|^2 \geq 0. \end{aligned}$$

Therefore, $G_i z^{l(i,k)} - x_i^k \to 0$ for $i = 1, \ldots, n$. □

Lemma 6. *Suppose assumptions 1–4 hold and that Algorithm 1 produces an infinite sequence of iterates with $\{x_i^k\}$ and $\{y_i^k\}$ being bounded. Then, for all $i \in \mathcal{I}_\text{B}$, one has $y_i^k - w_i^{l(i,k)} \to 0$.*

Proof. The argument to is siimlar to those of [14, Lemma 13] and [14, Theorem 1 (part 2)]: the crux of the proof is to establish for all $k \geq M$ that

$$\psi_k + \sum_{i\in\mathcal{I}_\text{F}}\langle x_i^k - G_i z^{l(i,k)}, T_i x_i^k - T_i G_i z^{l(i,k)}\rangle \geq (1-\sigma)\underline{\rho}\sum_{i\in\mathcal{I}_\text{B}}\|y_i^k - w_i^{l(i,k)}\|^2. \tag{21}$$

Since T_i is continuous and defined everywhere, x_i^k is bounded by assumption, and $z^{l(i,k)}$ is bounded by Lemma 2, the extreme value theorem implies that $T_i x_i^k - T_i G_i z^{l(i,k)}$ is bounded. Furthermore from Lemma 5, $\limsup_{k\to\infty}\psi_k \leq 0$, and $x_i^k - G_i z^{l(i,k)} \to 0$. Therefore the desired result follows from (21).

It remains to prove (21). For all $k \geq M$, we have

$$\begin{aligned} \psi_k &= \sum_{i=1}^n \langle G_i z^{l(i,k)} - x_i^k, y_i^k - w_i^{l(i,k)}\rangle \\ &\overset{\text{(a)}}{=} \sum_{i\in\mathcal{I}_\text{B}}\langle \rho_i^{l(i,k)}(y_i^k - w_i^{l(i,k)}) - e_i^{s(i,k)}, y_i^k - w_i^{l(i,k)}\rangle \\ &\quad + \sum_{i\in\mathcal{I}_\text{F}}\langle G_i z^{l(i,k)} - x_i^k, T_i G_i z^{l(i,k)} - w_i^{l(i,k)}\rangle \\ &\quad + \sum_{i\in\mathcal{I}_\text{F}}\langle G_i z^{l(i,k)} - x_i^k, y_i^k - T_i G_i z^{l(i,k)}\rangle \\ &\overset{\text{(b)}}{=} \sum_{i\in\mathcal{I}_\text{B}}\Big(\rho_i^{l(i,k)}\|y_i^k - w_i^{l(i,k)}\|^2 - \langle e_i^{s(i,k)}, y_i^k - w_i^{l(i,k)}\rangle\Big) \end{aligned}$$

$$
\begin{aligned}
&+\sum_{i\in\mathcal{I}_{\mathrm{F}}}\langle \rho_i^{l(i,k)}(T_iG_iz^{l(i,k)}-w_i^{l(i,k)}), T_iG_iz^{l(i,k)}-w_i^{l(i,k)}\rangle && (22)\\
&-\sum_{i\in\mathcal{I}_{\mathrm{F}}}\langle x_i^k-G_iz^{l(i,k)}, T_ix_i^k-T_iG_iz^{l(i,k)}\rangle\\
\overset{\text{(c)}}{\geq}\ & (1-\sigma)\underline{\rho}\sum_{i\in\mathcal{I}_{\mathrm{B}}}\|y_i^k-w_i^{l(i,k)}\|^2-\sum_{i\in\mathcal{I}_{\mathrm{F}}}\langle x_i^k-G_iz^{l(i,k)}, T_ix_i^k-T_iG_iz^{l(i,k)}\rangle.
\end{aligned}
$$

In the above derivation, (a) follows by substition of (8) into the $\mathcal{I}_{\mathrm{B}}$ terms and algebraic manipulation of the $\mathcal{I}_{\mathrm{F}}$ terms. Next, (b) is obtained by algebraic simplification of the $\mathcal{I}_{\mathrm{B}}$ terms and substitution of (9) into the two groups of $\mathcal{I}_{\mathrm{F}}$ terms. Finally, (c) follows by substituting the error criterion (11) from Assumption 3 into the $\mathcal{I}_{\mathrm{B}}$ terms and dropping the terms from (22), which must be nonnegative. □

Theorem 1. *Suppose assumptions 1–4 hold. If Algorithm 1 terminates at line 34, then its final iterate* $(z^{k+1}, \mathbf{w}^{k+1})$ *is a member of the extended solution set* $\mathcal{S}$ *defined in* (5)*. Otherwise, the sequence* $\{(z^k, \mathbf{w}^k)\}$ *generated by Algorithm 1 converges weakly to some point* $(\bar{z}, \overline{\mathbf{w}})$ *in* $\mathcal{S}$ *and furthermore* $x_i^k \rightharpoonup G_i\bar{z}$ *and* $y_i^k \rightharpoonup \overline{w}_i$ *for all* $i = 1, \ldots, n-1$*,* $x_n^k \rightharpoonup \bar{z}$*, and* $y_n^k \rightharpoonup -\sum_{i=1}^{n-1} G_i^*\overline{w}_i$*.*

Proof. The argument when the algorithm terminates via line 34 is identical to [14, Theorem 1]. From now on we assume the algorithm produces an infinite sequence of iterates. The proof proceeds by showing that the two conditions of Lemma 4 are satisfied. To establish Lemma 4(1) and Lemma 4(2) for $i \in \mathcal{I}_{\mathrm{B}}$, we will show that $\{x_i^k\}$ and $\{y_i^k\}$ are bounded, and then employ Lemmas 5 and 6. This argument is only a slight variation of what was given in [14]. The main departure from [14] is in establishing Lemma 4(2) for $i \in \mathcal{I}_{\mathrm{F}}$, which requires significant innovation.

We begin by establishing that $\{x_i^k\}$ and $\{y_i^k\}$ are bounded. For $i \in \mathcal{I}_{\mathrm{B}}$ the boundedness of $\{x_i^k\}$ follows exactly the same argument as [9, Lemma 10]. For $i \in \mathcal{I}_{\mathrm{F}}$ write using Lemma 1

$$
\begin{aligned}
\|x_i^k\| &\leq \|G_iz^{l(i,k)} - \hat{\rho}_i^{l(i,k)}T_iG_iz^{l(i,k)}\| + \hat{\rho}_i^{l(i,k)}\|w_i^{l(i,k)}\| && (23)\\
&\leq \|G_i\|\|z^{l(i,k)}\| + \overline{\rho}\|T_iG_iz^{l(i,k)}\| + \overline{\rho}\|w_i^{l(i,k)}\|. && (24)
\end{aligned}
$$

Now $z^{l(i,k)}$ and $w_i^{l(i,k)}$ are bounded by Lemma 2. Furthermore, since T_i is continuous with full domain, G_i is bounded, and $z^{l(i,k)}$ is bounded, $\{T_iG_iz^{l(i,k)}\}$ is bounded by the extreme value theorem. Thus $\{x_i^k\}$ is bounded for $i \in \mathcal{I}_{\mathrm{F}}$.

Now we prove that $\{y_i^k\}$ is bounded. For $i \in \mathcal{I}_{\mathrm{B}}$, Lemma 1 implies that

$$
y_i^k = \left(\rho_i^{l(i,k)}\right)^{-1}\left(G_iz^{l(i,k)} - x_i^k + \rho_i^{l(i,k)}w_i^{l(i,k)} + e_i^{s(i,k)}\right).
$$

Since ρ_i^k is bounded from above and below, G_i is bounded, $z^{l(i,k)}$ and $w_i^{l(i,k)}$ are bounded by Lemma 2, and $e_i^{s(i,k)}$ is assumed to be bounded, $\{y_i^k\}$ is bounded for $i \in \mathcal{I}_{\mathrm{B}}$. For $i \in \mathcal{I}_{\mathrm{F}}$, since $y_i^k = T_ix_i^k$ and T_i is continuous with full domain, it follows again from the extreme value theorem that $\{y_i^k\}$ is bounded.

Therefore we can apply Lemma 5 to infer that $G_i z^{l(i,k)} - x_i^k \to 0$ for $i = 1, \dots, n$, and Lemma 4(1) holds. Furthermore we can apply Lemma 6 to infer that $y_i^k - w_i^{l(i,k)} \to 0$ for $i \in \mathcal{I}_{\mathrm{B}}$.

It remains to establish that $y_i^k - w_i^{l(i,k)} \to 0$ for $i \in \mathcal{I}_{\mathrm{F}}$. The argument needs to be significantly expanded from that in [14], since it is not immediate that the stepsize $\hat{\rho}_i^k$ is bounded away from 0.

From Lemma 2, we know that $z^{l(i,k)}$ and $w_i^{l(i,k)}$ are bounded, as is the operator G_i by assumption. Furthermore, since T_i is continuous with full domain, we know once again from the extreme value theorem that there exists $B \geq 0$ such that

$$(\forall k \in \mathbb{N}) \quad \|T_i G_i z^{l(i,k)} - w_i^{l(i,k)}\| \leq B. \tag{25}$$

We have already shown that x_i^k is bounded. Using the boundedness of z^k and w_i^k in conjunction with Assumption 4 and inspecting the steps in the backtracking search, there must exist a closed ball $\mathcal{B}_x \subset \mathcal{H}_i$ such that $\tilde{x}_i^{(j,s(i,k))} \in \mathcal{B}_x$ for all $k, j \in \mathbb{N}$ such that $i \in I_k$ and j is encountered during the backtracking linesearch at step k. In addition, let $\mathcal{B}_{GZ} \subset \mathcal{H}_i$ be a closed ball containing $G_i z^{l(i,k)}$ for all $k \in \mathbb{N}$. Let $\mathcal{B} = \mathcal{B}_x \cup \mathcal{B}_{GZ}$, which is another closed ball. Since $\mathcal{H}_i$ is finite dimensional, $\mathcal{B}$ is compact. Since T_i is continous everywhere, by the Heine-Cantor theorem it is uniformly continuous on $\mathcal{B}$ [19, Theorem 21.4].

Continuing, we write

$$y_i^k - w_i^{l(i,k)} = T_i x_i^k - w_i^{l(i,k)} = T_i G_i z^{l(i,k)} - w_i^{l(i,k)} + T_i x_i^k - T_i G_i z^{l(i,k)}. \tag{26}$$

Since T_i is uniformly continuous on $\mathcal{B}$ it must be Cauchy continuous, meaning that $x_i^k - G_i z^{l(i,k)} \to 0$ implies $T_i x_i^k - T_i G_i z^{l(i,k)} \to 0$. Thus, to prove that $y_i^k - w_i^{l(i,k)} \to 0$ it is sufficient to show that $T_i G_i z^{l(i,k)} - w_i^{l(i,k)} \to 0$.

We now show that indeed $T_i G_i z^{l(i,k)} - w_i^{l(i,k)} \to 0$. Fix $\epsilon > 0$. Since T_i is uniformly continuous on $\mathcal{B}$, there exists $\delta > 0$ such that whenever $x, y \in \mathcal{B}$ and $\|x - y\| \leq \delta$, then $\|T_i x - T_i y\| \leq \epsilon/4$. Since $G_i z^{l(i,k)} - x_i^k \to 0$, there exists $K \geq 1$ such that for all $k \geq K$,

$$\|G_i z^{l(i,k)} - x_i^k\| \leq \epsilon \min\left(\frac{\nu\epsilon}{4B\Delta}, \frac{\nu\delta}{B}, \underline{\rho}\right) \tag{27}$$

with B as in (25), Δ from the linesearch termination criterion, and $\underline{\rho}$ from Assumption 4. For any $k \geq K$ we will show that

$$\|T_i G_i z^{l(i,k)} - w_i^{l(i,k)}\| \leq \epsilon. \tag{28}$$

If $\|T_i G_i z^{l(i,k)} - w_i^{l(i,k)}\| \leq \epsilon/2$, then (28) clearly holds. So from now on it is sufficient to consider k for which $\|T_i G_i z^{l(i,k)} - w_i^{l(i,k)}\| > \epsilon/2$. Again, let $\rho_i^{(j,s(i,k))} = \rho$ for brevity. Reconsidering (14), we now have the following lower bound for the left-hand side of (12):

$$\frac{\|T_i G_i z^{l(i,k)} - w_i^{l(i,k)}\| - \left\|T_i\big(G_i z^{l(i,k)} - \rho(T_i G_i z^{l(i,k)} - w_i^{l(i,k)})\big) - T_i G_i z^{l(i,k)}\right\|}{\rho\|T_i G_i z^{l(i,k)} - w_i^{l(i,k)}\|}$$

$$> \frac{\epsilon/2 - \left\|T_i\big(G_i z^{l(i,k)} - \rho(T_i G_i z^{l(i,k)} - w_i^{l(i,k)})\big) - T_i G_i z^{l(i,k)}\right\|}{\rho\|T_i G_i z^{l(i,k)} - w_i^{l(i,k)}\|}. \tag{29}$$

Now, suppose it were true that

$$\|G_i z^{l(i,k)} - \tilde{x}_i^{(j,s(i,k))}\| = \rho\|T_i G_i z^{l(i,k)} - w_i^{l(i,k)}\| \leq \delta. \tag{30}$$

Then the uniform continuity of T_i on $\mathcal{B}$ would imply that

$$\|T_i G_i z^{l(i,k)} - T_i \tilde{x}_i^{(j,s(i,k))}\| = \left\|T_i\big(G_i z^{l(i,k)} - \rho(T_i G_i z^{l(i,k)} - w_i^{l(i,k)})\big) - T_i G_i z^{l(i,k)}\right\| \leq \frac{\epsilon}{4}.$$

We next observe that (30) is implied by $\rho \leq \frac{\delta}{B}$, in which case (29) gives the following lower bound on the left-hand side of (12):

$$\frac{\langle G_i z^{l(i,k)} - \tilde{x}_i^{(j,s(i,k))}, \tilde{y}_i^{(j,s(i,k))} - w_i^{l(i,k)}\rangle}{\|G_i z^{l(i,k)} - \tilde{x}_i^{(j,s(i,k))}\|^2} > \frac{\epsilon}{4\rho\|T_i G_i z^{l(i,k)} - w_i^{l(i,k)}\|} \geq \frac{\epsilon}{4\rho B}.$$

Therefore if ρ also satisfies $\rho \leq \frac{\epsilon}{4B\Delta}$, then

$$\frac{\langle G_i z^{l(i,k)} - \tilde{x}_i^{(j,s(i,k))}, \tilde{y}_i^{(j,s(i,k))} - w_i^{l(i,k)}\rangle}{\|G_i z^{l(i,k)} - \tilde{x}_i^{(j,s(i,k))}\|^2} > \Delta. \tag{31}$$

Thus, any stepsize satisfying $\rho \leq (1/B)\min\{\epsilon/4\Delta, \delta\}$ must cause the backtracking linesearch termination criterion at line 18 to hold. Therefore, since the backtracking linesearch proceeds by reducing the stepsize by a factor of ν at each inner iteration, it must terminate with

$$\hat{\rho}_i^{l(i,k)} \geq \underline{\rho}^{bt} \triangleq \min\left\{\frac{\nu\epsilon}{4B\Delta}, \frac{\nu\delta}{B}, \underline{\rho}\right\}. \tag{32}$$

Now, using Lemma 1, we have

$$x_i^k - G_i z^{l(i,k)} = -\hat{\rho}_i^{l(i,k)}(T_i G_i z^{l(i,k)} - w_i^{l(i,k)})$$

and therefore

$$\|x_i^k - G_i z^{l(i,k)}\| = \hat{\rho}_i^{l(i,k)}\|T_i G_i z^{l(i,k)} - w_i^{l(i,k)}\|.$$

Thus,

$$\begin{aligned}\|T_i G_i z^{l(i,k)} - w_i^{l(i,k)}\| &\leq (\underline{\rho}^{bt})^{-1}\|x_i^k - G_i z^{l(i,k)}\| \\ &\leq \min\left\{\frac{\nu\epsilon}{4B\Delta}, \frac{\nu\delta}{B}, \underline{\rho}\right\}^{-1}\|x_i^k - G_i z^{l(i,k)}\| \leq \epsilon.\end{aligned}$$

and therefore (28) holds for all $k \geq K$. Since $\epsilon > 0$ was chosen arbitrarily, it follows that $\|T_i G_i z^{l(i,k)} - w_i^{l(i,k)}\| \to 0$ and thus $\|y_i^k - w_i^{l(i,k)}\| \to 0$ by (26). The proof that Lemma 4(2) holds is now complete. The proof of the theorem now follows from Lemma 4. □

If $\mathcal{H}_i$ is not finite dimensional for $i \in \mathcal{I}_{\mathrm{F}}$, Theorem 1 can still be proved if the assumption on T_i is strengthened to Cauchy continuity over all bounded sequences. This is slightly stronger than the assumption given in [22, Equation (1.1)] for proving weak convergence of Tseng's forward-backward-forward method. That assumption is Cauchy continuity but only for all *weakly convergent* sequences.

References

[1] Alotaibi, A., Combettes, P.L., Shahzad, N.: Solving Coupled Composite Monotone Inclusions by Successive Fejér Approximations of their Kuhn–Tucker Set. SIAM Journal on Optimization **24**(4), 2076–2095 (2014)

[2] Bauschke, H.H., Combettes, P.L.: Convex analysis and monotone operator theory in Hilbert spaces. Springer Science & Business Media (2011)

[3] Bello Cruz, J., Díaz Millán, R.: A variant of forward-backward splitting method for the sum of two monotone operators with a new search strategy. Optimization **64**(7), 1471–1486 (2015)

[4] Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine Learning **3**(1), 1–122 (2011)

[5] Combettes, P.L., Eckstein, J.: Asynchronous block-iterative primal-dual decomposition methods for monotone inclusions. Mathematical Programming **168**(1-2), 645–672 (2018)

[6] Combettes, P.L., Pesquet, J.C.: Proximal splitting methods in signal processing. In: Fixed-point algorithms for inverse problems in science and engineering, pp. 185–212. Springer (2011)

[7] Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward-backward splitting. Multiscale Modeling & Simulation **4**(4), 1168–1200 (2005)

[8] Davis, D., Yin, W.: A three-operator splitting scheme and its optimization applications. Set-Valued and Variational Analysis **25**(4), 829–858 (2017)

[9] Eckstein, J.: A simplified form of block-iterative operator splitting and an asynchronous algorithm resembling the multi-block alternating direction method of multipliers. Journal of Optimization Theory and Applications **173**(1), 155–182 (2017)

[10] Eckstein, J., Svaiter, B.F.: A family of projective splitting methods for the sum of two maximal monotone operators. Mathematical Programming **111**(1), 173–199 (2008)

[11] Eckstein, J., Svaiter, B.F.: General projective splitting methods for sums of maximal monotone operators. SIAM Journal on Control and Optimization **48**(2), 787–811 (2009)

[12] Iusem, A., Svaiter, B.: A variant of Korpelevich's method for variational inequalities with a new search strategy. Optimization **42**(4), 309–321 (1997)

[13] Johnstone, P.R., Eckstein, J.: Convergence rates for projective splitting. arXiv preprint arXiv:1806.03920 (2018)

[14] Johnstone, P.R., Eckstein, J.: Projective splitting with forward steps: Asynchronous and block-iterative operator splitting. arXiv preprint arXiv:1803.07043 (2018)

[15] Korpelevich, G.: Extragradient method for finding saddle points and other problems. Matekon **13**(4), 35–49 (1977)

[16] Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM Journal on Numerical Analysis **16**(6), 964–979 (1979)

[17] Malitsky, Y., Tam, M.K.: A forward-backward splitting method for monotone inclusions without cocoercivity. arXiv preprint arXiv:1808.04162 (2018)

[18] Mercier, B., Vijayasundaram, G.: Lectures on topics in finite element solution of elliptic problems. Tata Institute of Fundamental Research, Bombay (1979)

[19] Ross, K.A.: Elementary analysis: The Theory of Calculus. Springer (1980)

[20] Solodov, M.V., Svaiter, B.F.: A hybrid approximate extragradient–proximal point algorithm using the enlargement of a maximal monotone operator. Set-Valued Analysis **7**(4), 323–345 (1999)

[21] Tran-Dinh, Q., Vũ, B.C.: A new splitting method for solving composite monotone inclusions involving parallel-sum operators. Preprint 1505.07946, arXiv (2015)

[22] Tseng, P.: A modified forward-backward splitting method for maximal monotone mappings. SIAM Journal on Control and Optimization **38**(2), 431–446 (2000)